Case study

リード化合物の研究を効率化する ドラッグデザインの自動化に挑む

HP ProLiant DL980 G7上の広大なメモリー空間にリード化合物創出研究のための多様なデータを展開し、検索から評価、そして再検索というサイクルを自動的に回す

業界
医療・製薬

目的
データベース検索に基づく化合物デザインの自動化

アプローチ
- 大容量のメモリーと可能な限り多くのコア数が搭載できるDBサーバーを導入
- 高性能なインテル® Xeon® プロセッサー E7ファミリーを8基（80コア）搭載したHP ProLiant DL980 G7を導入
- データベースI/Oの高速化を狙い、内蔵ストレージにHP PCIe IOアクセラレータ for ProLiantサーバーを採用

ITの効果
- 検索時間の大幅な効率化とスピードアップを図る
- HP通報サービスの活用で、研究と並行したシステム運用管理を実現

ビジネスの効果
- 高い可能性を期待できるリード化合物の発見につなげる
- 医薬品開発の期間短縮と効率化を目指す

astellas

データベース

従来、検索をかけるための検索式の作成、得られた検索結果をシミュレーションした結果の評価、といった工程は研究者が手作業で行っていました。こうした工程を、機械学習の技術も取り入れながら自動化し、検索のサイクルを網羅的に数多く回すことができれば、よりリード化合物の合理的なデザインが可能になるはずです。そのために、検索式の自動生成、シミュレーションシステムへの検索結果の投入、シミュレーション結果の評価、これを反映した検索式の修正といった作業を、DBサーバーが司令塔のように自動的にこなす処理フローを実現したいと構想していました。

片山直子氏　アステラス製薬株式会社 研究本部 化学研究所 主管研究員

新薬の研究開発にかかる時間とコストは年々、拡大を続けている。同時に、医薬品の候補物質となることが期待できる「リード化合物」の探索をランダムに行うことは極めて効率が悪い。売上規模で国内2位、世界でもトップ20以内にランクされるグローバルな製薬会社、アステラス製薬でも、業界共通の悩みであるこうした課題を克服すべく、近年のゲノミクスやプロテオミクスなどの研究で解明されてきた知識や情報、さらに自らの合成成果や社内の実験結果データを集積し、リード化合物の探索や効果の改善にコンピューターを積極的に活用している。偶然に頼った“セレンディピティ”から、理論と合理性に基づく“ラショナル”へ。その挑戦的な取り組みをHP ProLiant DL980 G7が支えている。

アステラス製薬株式会社 研究本部 化学研究所
主管研究員
片山 直子 氏

アステラス製薬株式会社 研究本部 化学研究所
藤 秀義 氏

ゲノミクスなどの研究の進展により医薬品開発でも計算機が重責を担う

それまでに存在しなかった新しい医薬品を世の中に送り出すまでには、10年以上にわたる長期の研究開発作業と莫大なコストがかかるものだ。こうした取り組みを効率化し、開発のスピードアップやコストの削減を図るため、特に“宝探し”的な要素が強い医薬品開発の初期段階において、医薬品の候補物質となることが期待できる「リード化合物」の特定や効果の改善に、近年コンピューターが積極的に活用されている。

「リード化合物」とは、新薬の開発を目指す病気と関連性が深い標的、主にタンパク質に何らかの作用を及ぼす化合物を指す。人間に対する毒性などについてはまだ考慮しない。ゲノミクスやプロテオミクスといった分野の研究が急速に進み、遺伝子やこれと関連するタンパク質、ある病気に関係するタンパク質、さらにはタンパク質自体の構造など様々な情報が膨大に集積。これにより、リード化合物を探し出す、あるいは効果がさらに高まるようにリード化合物の構造をデザインする、といった作業にコンピューターの適用が可能になったという背景がある。

医療用医薬品を中核事業として売上規模で国内2位、世界でもトップ20以内にランクされ、グローバルな舞台で活躍する有数の研究開発型製薬会社、アステラス製薬。「先端・信頼の医薬で、世界の人々の健康に貢献する」という経営理念を掲げる同社も、こうしたリード化合物の探索やデザインに早くからコンピューターを活用してきた。その中心的役割を担うのが化学研究所である。

同研究所では、長年取り組んできたリード化合物研究のさらなる効率化とスピードアップを図るため、より高い効果が見込めるリード化合物のデザインプロセスを自動化しようとしている。そのためのシステムで中核となるデータベースサーバー（DBサーバー）として採用されたのが、超高速なI/Oを誇るHP PCIe IOアクセラレータ for ProLiantサーバー（以下、HP IOアクセラレータ）を内蔵した、高性能なインテル® Xeon® プロセッサー E7ファミリーを8基（80コア）搭載のHP ProLiant DL980 G7だった。

リード化合物の研究をセレンディピティからラショナルに

「ヒトゲノムやその情報を基に体内で作られるタンパク質などに関する広範な知識が集積される以前から、リード化合物の探索や機能強化は医薬品開発に不可欠なステージです。しかし、その手法は研究者の経験や勘と膨大な回数の実験を繰り返すことで、偶然にたどり着く“セレンディピティ”なものでした。現在では、遺伝子やタンパク質、さらには塩基配列を超えた遺伝子の振る舞いを対象としたエピジェネティクスなどの研究から多くの知識が得られています。これらを基に、標的とするタンパク質などの対象を特定して、これを狙った医薬品に向けてリード化合物を合理的に、意識的に開発する“ラショナル”な取り組みが必要です」と、同研究所で主管研究員を務める片山直子氏は、今日的な研究のトレンドを解説する。

こうした“ラショナル” さを追求するため、同研究所では、最新の手法であるFBDD（Fragment Based Drug Design）を用いて研究を進めている。「当研究所は、私たちの計算化学チームに加え、探索や機能強化のためにリード化合物の多様な誘導体を一気に合成するコンビナトリアル合成チーム、合成したリード候補化合物の標的タンパク質への結合状態を計測するX線結晶構造解析チームで構成。FBDDに必要な三つの要素を結集した体制を作り上げています。特に、自ら合成した化合物の特性を計測し、独自のデータにできるX線結晶構造解析チームを有しているのはアステラス製薬の大きな強みであり、年間で1000件以上の構造解明をしています」（片山氏）。

こうした化学研究所の実力は社会貢献の面でも遺憾なく発揮されている。国際的な取り組みが進む「顧みられない熱帯感染症（NTDs）」の創薬研究のために世界で初めて構築・公開された統合データベース「iNTRODB（イントロ・ディービー）」の開発に参画。2013年の産官学連携功労者表彰で厚生労働大臣賞を受賞するなど、社会的にも高く評価されている。

検索のサイクルを数多く回し有効なリード化合物の発見につなげる

今回、新たなチャレンジのスタートを切ろうとしている、高性能DBサーバー導入によるリード化合物デザインの自動化も、同研究所が進めてきた“セレンディピティ”から“ラショナル”へという取り組みの一環と見ることができる。

リード化合物をデザインするプロセスは次のようなものだ。まず、化合物やタンパク質、化合物とタンパク質の複合体などに関する多様なデータを社内や公開データベースから収集して統合データベースを構築。このデータベースに検

■DBサーバーの概要

インテル® Xeon® プロセッサー
E7ファミリー

ハードウェア: HP ProLiant DL980 G7×1台
搭載プロセッサー:インテル® Xeon® プロセッサー E7-4870 2.4GHz 8基／80コア
搭載メモリー:1TB
オプション:HP PCIe IOアクセラレータ for ProLiantサーバー 2TB×6台
OS:Red Hat Enterprise Linux

索をかけ、リード化合物としてより効果が高そうであったり、構造の一部を変えることで効果が高まりそうであったりする化合物を抽出。得られた結果を分子動力学計算、ドッキングシミュレーションといったシミュレーションで評価し、その結果を基に再度検索をかけるというサイクルを繰り返す。そのサイクル回数が多いほど、リード化合物として可能性の高いものを選び出すことができる、つまりはデザインできるということになる。

「従来、検索をかけるための検索式の作成、得られた検索結果をシミュレーションした結果の評価、といった工程は研究者が手作業で行っていました。このため、研究者が想定していない候補化合物の取りこぼしなども起きていただろうと考えていました。こうした工程を、機械学習の技術も取り入れながら自動化し、検索のサイクルを網羅的に数多く回すことができれば、より合理的なデザインが可能になるはずです。そのために、検索式の自動生成、クラスター構成で整備してきたシミュレーションシステムへの検索結果の投入、シミュレーション結果の評価、その評価を反映した検索式の修正といった作業を、DBサーバーが司令塔のように自動的にこなす処理フローを実現したいという構想を持っていたのです」と片山氏は振り返る。

大量のメモリーとコア数が使えることをサーバー選定の基準に

片山氏とともに新しいDBサーバーの機種選定にあたった同研究所の藤 秀義氏は、必要なサーバーの要件として、できるだけ大きなメモリー量を使用できること、コア数も可能な限り大きくできること、という2点を重視していた。

「検索のスピードを上げるには、データベースの全データを一筐体内のメモリー上に展開しておく方が有利です。これがまず可能なこと。そして、検索自体は複数の処理を同時並列で流したいと考えていたため、できるだけ多くのコア数を利用できるサーバーを、と考えていたのです」（藤氏）。

実は、化学研究所は長年のHP ProLiantサーバーユーザーである。このため、今回の新DBサーバー導入を検討する際、すでに導入していた40コアのHP ProLiant DL580 G7上で、想定していたデザイン自動化のサイクルをきちんと回せるかどうか、小規模なテスト用データを使って事前に検証していた。

「そのときの感触は悪くありませんでした。HP ProLiantのラインには、DL580の上に、CPUを80コア搭載でき、メモリーも最大4TBまで拡張できるHP ProLiant DL980 G7があることを聞いていました。これだけのコア数とメモリー量が使えれば、別の新しいチャレンジもできそうだと感じたのです。その際には、データベースへの書き込みや更新といった処理が発生する可能性もあります。この点で、非常に高速なストレージI/Oを提供するHP IOアクセラレータが内蔵できることも大きな魅力でした。もちろん、ストレージI/Oの高速化はデザイン自動化のさらなるスピードアップにも貢献してくれるだろうという感触もありました」（藤氏）。

HP通報サービスの活用により本来の研究業務に集中

また、性能以外の面でもHP ProLiantサーバーには大きな魅力があった、と藤氏は付け加える。「サーバーの運用や管理は研究所内の計算化学チームが担当しています。もちろん、本来の研究活動をしながらの対応になるため、できるだけメンテナンスや障害対応といったところに手間を取られたくないというのが本音です。HP ProLiantサーバーをこれまで使ってきた中で、HP通報サービスには非常にお世話になっていました。マシンルームに出向かなくても、HPから電話やメールを介して、サーバーで発生して

いるトラブルなどを知らせてくれるのです。非常に安心だし、運用管理負担を大幅に軽くしてくれています。一度利用してしまうとやめられないサービスです。これもHP ProLiant DL980 G7を選んだ大きな理由の一つでした」と、藤氏は笑顔を見せる。

2013年9月中旬に納入されたHP ProLiant DL980 G7には、OSとしてRed Hat Enterprise Linux、データベースソフトとしてPostgreSQLがインストールされ、少しずつテスト稼働に向けた準備が進んでいる。社内のデータを集約したOracleデータベースとの接続も始まろうとしている。「ストレージを除けば、基本的に事前検証で使ったHP ProLiant DL580 G7の構成と同じであるため、移行に関わる特別な負担もなさそうです」（藤氏）。

バーチャルライブラリーの高速作成、PPIなどの領域でも新サーバーに期待

HP ProLiant DL980 G7を核にしたリード化合物のデザイン自動化システムは、今後、搭載したデータベースの情報量を確実に増やしながら、徐々に構想の実現に向けて歩み出している。この新しいDBサーバーは、デザイン自動化にとどまらない先進的な研究へのチャレンジにも大きな効果をもたらしてくれるものと、同研究所のメンバーたちは期待を寄せる。

「リード化合物をどのように修飾したり、構造の一部を変換したりするかを検討する際、その可能性を探るためにコンピューター上で膨大な数の化合物をバーチャルライブラリーとして作成することがあります。こうした作業も、大量のコア数とメモリーが使えることで、大幅なスピードアップが図れると見ています」（藤氏）。

「リード化合物の探索として、今後取り組む必要があるのは、単純に一つのタンパク質を標的にしたものではなく、複数のタンパク質が関係するタンパク質間相互作用（PPI）のようなより複雑な系への取り組みです。対象となるタンパク質のデータサイズも大きくなるうえ、結合部位も複数を考慮しなくてはなりません。こうした新しい領域の研究に、大量のコア数とメモリーを活用できる意味は大きいはずです。また、リード化合物を作り出した次には、化合物の毒性や体内での代謝などを研究するリード化合物最適化のステップが待っています。そこに計算化学の技術を適用していく道を探ることも重要なテーマ。新たなチャレンジングな取り組みでも、HPには支援を期待しています」と、片山氏は話を締めくくった。

ソリューション概略

導入ハードウェア
HP ProLiant DL980 G7
HP PCIe IOアクセラレータ for ProLiantサーバー

安全に関するご注意　ご使用の際は、商品に添付の取扱説明書をよくお読みの上、正しくお使いください。水、湿気、油煙等の多い場所に設置しないでください。火災、故障、感電などの原因となることがあります。

お問い合わせはカスタマー・インフォメーションセンターへ
03-5749-8328 月～金 9:00～19:00　土 10:00～17:00（日、祝祭日、年末年始および5/1を除く）
機器のお見積もりについては、代理店、または弊社営業にご相談ください。
HP ProLiantに関する情報は http://www.hp.com/jp/proliant

Intel、インテル、Intel ロゴ、Intel Inside、Intel Inside ロゴ、Xeon、Xeon Insideは、アメリカ合衆国および / またはその他の国におけるIntel Corporationの商標です。
記載されている会社名および商品名は、各社の商標または登録商標です。
記載事項は2013年12月現在のものです。
本カタログに記載されている情報は取材時におけるものであり、閲覧される時点で変更されている可能性があります。あらかじめご了承ください。
© Copyright 2013 Hewlett-Packard Development Company,L.P.

本カタログは、環境に配慮した用紙と植物性大豆油インキを使用しています。 PRINTED WITH SOY INK

日本ヒューレット・パッカード株式会社
〒136-8711 東京都江東区大島2-2-1